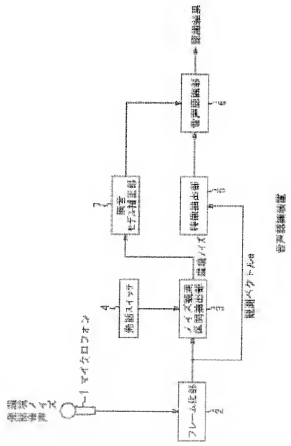
DEVICE AND METHOD FOR MODEL ADAPTATION, RECORDING MEDIUM, AND VOICE RECOGNITION DEVICE

Publication number: JP2002156992 (A)
Publication date: 2002-05-31
Inventor(s): NAKATSUKA KOUCHIYO +
Applicant(s): SONY CORP +
Classification:
- international: ***G10L15/06; G10L15/20;*** **G10L15/00;** (IPC1-7): G10L15/06; G10L15/20
- European:
Application number: JP20000353790 20001121
Priority number(s): JP20000353790 20001121

Abstract of **JP 2002156992 (A)**

PROBLEM TO BE SOLVED: To adapt a soundless model which is sufficiently adaptive to a soundless part in a voice section. SOLUTION: A noise observation section extraction part 3 extracts environment noise observed in the section right before a sound section and supplies it to a soundless model correction part 7. The soundless model correction part 7 adapts a soundless model as a sound model showing no sound according to the environment noise in the section right before the sound section of a current voice to be recognized and environment noise in the section right before a sound section of a voice as an object of past voice recognition.



(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号
特開2002－156992
（P2002－156992A）
(43)公開日　平成14年５月31日(2002.5.31)

(51)Int.Cl.⁷	識別記号	ＦＩ		テーマコード（参考）
Ｇ１０Ｌ　15/06		Ｇ１０Ｌ　3/00	５２１Ｔ	５Ｄ０１５
15/20			５３１Ｑ	

審査請求　未請求　請求項の数11　ＯＬ　（全 23 頁）

(21)出願番号　特願2000－353790(P2000－353790)

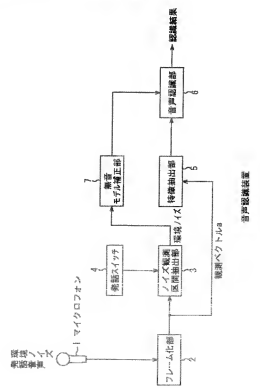
(22)出願日　平成12年11月21日(2000.11.21)

(71)出願人　000002185
ソニー株式会社
東京都品川区北品川６丁目７番35号
(72)発明者　中塚　洪長
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(74)代理人　100082131
弁理士　稲本　義雄

Ｆターム(参考)　5D015 EE05 GG00

(54)【発明の名称】　モデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置

(57)【要約】
【課題】　音声区間中の無音部分に十分対処可能な無音モデルの適応を行う。
【解決手段】　ノイズ観測区間抽出部３は、音声区間の直前の区間で観測される環境ノイズを抽出し、無音モデル補正部７に供給する。無音モデル補正部７は、現在の音声認識の対象となっている音声の音声区間の直前の区間における環境ノイズと、過去に音声認識の対象とされた音声の音声区間の直前の区間における環境ノイズとに基づいて、無音を表す音響モデルである無音モデルの適応を行う。



【特許請求の範囲】

【請求項１】　音声を認識するのに用いる音響モデルの適応を行うモデル適応装置であって、
音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出手段と、
現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応手段とを備えることを特徴とするモデル適応装置。

【請求項２】　前記モデル適応手段は、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた１以上の音声の音声区間の直前の区間における抽出データとから、現在の音声認識の対象となっている音声の認識に用いる前記無音モデルを生成することを特徴とする請求項１に記載のモデル適応装置。

【請求項３】　前記モデル適応手段は、過去に音声認識の対象とされた１以上の音声の音声区間の直前の区間における抽出データに基づいて、第１の無音モデルを生成するとともに、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データに基づいて、第２の無音モデルを生成し、前記第１と第２の無音モデルに基づいて、現在の音声認識の対象となっている音声の認識に用いる前記無音モデルを生成することを特徴とする請求項１に記載のモデル適応装置。

【請求項４】　前記モデル適応手段は、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データに基づいて、前記無音モデルを生成し、その無音モデルと、過去に音声認識の対象とされた音声の認識に用いられた前記無音モデルとに基づいて、現在の音声認識の対象となっている音声の認識に用いる前記無音モデルを生成することを特徴とする請求項１に記載のモデル適応装置。

【請求項５】　前記音声の認識は、音声の特徴空間における特徴量のベクトルまたは特徴量の分布に基づいて行われ、
前記モデル適応手段は、前記抽出データから得られる前記特徴量のベクトルまたは特徴量の分布に基づいて、前記無音モデルの適応を行うことを特徴とする請求項１に記載のモデル適応装置。

【請求項６】　前記モデル適応手段は、前記抽出データから得られる前記特徴量と特徴量の分布の両方に基づいて、前記無音モデルの適応を行うことを特徴とする請求項５に記載のモデル適応装置。

【請求項７】　前記モデル適応手段は、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データそれぞれに重みを付して、前記無音モデルの適応を行うことを特徴とする請求項１に記載のモデル適応装置。

【請求項８】　前記モデル適応手段は、統計的手法によって、前記無音モデルの適応を行うことを特徴とする請求項１に記載のモデル適応装置。

【請求項９】　音声を認識するのに用いる音響モデルの適応を行うモデル適応方法であって、
音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出ステップと、
現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応ステップとを備えることを特徴とするモデル適応方法。

【請求項１０】　音声を認識するのに用いる音響モデルの適応を行うモデル適応処理を、コンピュータに行わせるプログラムが記録されている記録媒体であって、
音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出ステップと、
現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応ステップとを備えるプログラムが記録されていることを特徴とする記録媒体。

【請求項１１】　音声を、音響モデルを用いて認識する音声認識装置であって、
音声データの特徴量を抽出する特徴抽出手段と、
前記特徴量と音響モデルに基づいて、前記音声を認識する音声認識手段と、
音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出手段と、
現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応手段とを備えることを特徴とする音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、モデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置に関し、特に、例えば、ノイズに起因する音声認識性能の劣化を防止することができるようにするモデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置に関する。

【０００２】

【従来の技術】音声認識装置においては、例えば、音声認識対象の音声から、その特徴ベクトルが抽出され、そ

の特徴ベクトルの系列が、音声の音響モデルから観測される尤度を計算すること等によって、音声が認識される。

【０００３】

【発明が解決しようとする課題】ところで、音声認識装置においては、一般に、ユーザの発話が行われている区間である音声区間を特定し、その音声区間を対象に、音声認識が行われる。

【０００４】しかしながら、ユーザの音声は、音声区間の全体にわたって存在するとは限らない。即ち、音声区間には、一般に、息継ぎ等によって、ユーザの音声が存在しない部分がある。

【０００５】一方、音声認識装置が使用される環境においては、認識対象の音声以外の音、即ち、ノイズが存在する。

【０００６】具体的には、例えば、音声を入力するマイク（マイクロフォン）を叩く音や、場所によっては、ドアを開閉する音、ユーザの咳の音、音声認識しようとしている音声のユーザ以外のユーザの発話等が、ノイズとして存在する。また、例えば、音声認識装置が、エンターテイメント用のロボット等に適用された場合には、そのロボットに各種の動作を行わせるためのアクチュエータの音が、ノイズとして存在し、さらに、そのロボットが、デモンストレーション会場で公表されるときには、観衆の話し声や拍手等が、ノイズとして存在する。

【０００７】従って、音声区間において、ユーザの音声が存在しない部分には、上述したようなノイズのみが存在することとなるが、音声認識装置では、そのノイズのみの部分についても、ユーザの音声が存在するものとして、音響モデルを用いて、音声認識が行われるため、認識性能が劣化することがあった。即ち、特に、音声区間の開始から、実際に、ユーザの発話が開始されるまでの時間が長くなると、認識性能が低下する課題があった。

【０００８】そこで、ユーザの音声が存在しない状態、即ち、音声認識装置が使用される環境においてノイズが存在する場合には、そのノイズのみが存在する状態としての無音を表す音響モデルである無音モデルを導入し、音声区間の中で、ユーザの音声が存在しない部分（以下、適宜、無音部分という）については、その無音モデルで対処する方法がある。

【０００９】しかしながら、音声認識装置が使用される環境におけるノイズは、一定であるとは限らず、むしろ時々刻々と変化することが多いため、あらかじめ作成しておいた無音モデルを、そのまま用いるのでは、音声区間中の無音部分について、十分に対処することができない場合がある。

【００１０】そこで、本件出願人は、例えば、特開2000-259198号公報（特願平11-57467号）において、音声区間の直前の区間における音声（ノイズ）に基づいて、無音モデルの適応を行う方法について、先に提案している。

【００１１】しかしながら、先に提案した方法では、現在の音声認識の対象となっている音声の音声区間（以下、適宜、注目音声区間という）の直前の区間における音声にのみ基づいて、無音モデルの適応を行うため、例えば、注目音声区間の直前において、ユーザが、音声の入力に用いるマイクを叩く等した場合や、観衆が拍手を行った場合等の、いわば突発的なノイズが生じた場合、その突発的なノイズに基づいて、無音モデルの適応が行われることがあり、この場合、音声区間中の無音部分について、十分に対処することが困難であると考えられる。

【００１２】また、そのような突発的なノイズが生じず、比較的定常的なノイズが長時間連続している場合には、注目音声区間のみならず、過去に音声認識の対象とされた音声の音声区間の直前の区間におけるノイズをも用いて、無音モデルの適応を行った方が、音声区間中の無音部分について、より十分に対処することができると予想される。

【００１３】本発明は、このような状況に鑑みてなされたものであり、音声区間中の無音部分に十分対処可能な無音モデルの適応を行うことができるようにし、これにより、無音部分に起因する音声認識性能の劣化を防止（低減）することができるようにするものである。

【００１４】

【課題を解決するための手段】本発明のモデル適応装置は、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出手段と、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応手段とを備えることを特徴とする。

【００１５】本発明のモデル適応方法は、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出ステップと、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応ステップとを備えることを特徴とする。

【００１６】本発明の記録媒体は、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出ステップと、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応ステップとを備えるプログラムが記録されていこ

とを特徴とする。

【００１７】本発明の音声認識装置は、音声区間の直前の区間で観測される音声データを抽出し、抽出データとして出力するデータ抽出手段と、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応を行うモデル適応手段とを備えることを特徴とする。

【００１８】本発明のモデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置においては、音声区間の直前の区間で観測される音声データが抽出され、抽出データとして出力される。そして、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応が行われる。

【００１９】

【発明の実施の形態】図１は、本発明を適用した音声認識装置の一実施の形態の構成例を示している。

【００２０】この音声認識装置において、マイク１は、認識対象である発話音声を、環境ノイズとともに集音し、フレーム化部２に出力する。フレーム化部２は、マイク１から入力される音声データを、所定の時間間隔（例えば、１０ｍｓ）で取り出し、その取り出したデータを、１フレームのデータとして出力する。フレーム化部２が出力する１フレーム単位の音声データは、そのフレームを構成する時系列の音声データそれぞれをコンポーネントとする観測ベクトルａとして、ノイズ観測区間抽出部３、および特徴抽出部５に供給される。

【００２１】ここで、以下、適宜、第ｔフレームの音声データである観測ベクトルを、ａ（ｔ）とあす。

【００２２】ノイズ観測区間抽出部３は、フレーム化部２から入力されるフレーム単位の音声データを所定の時間（Ｍフレーム分以上）だけバッファリングし、図２に示すように、発話スイッチ４がオンとされるタイミングt_0からＭフレーム分だけ以前のタイミングt_1までをノイズ観測区間Ｔｎとして、そのノイズ観測区間ＴｎにおけるＭフレーム分の観測ベクトルａを抽出して、無音モデル補正部７に出力する。

【００２３】発話スイッチ４は、ユーザが発話を開始するときにユーザによってオンとされ、発話を終了するときにオフとされる。したがって、発話スイッチ４がオンとされたタイミングt_0以前（ノイズ観測区間Ｔｎ）の音声データには、発話音声は含まれず、環境ノイズだけが存在する。また、発話スイッチ４がオンとされたタイミングt_0から発話スイッチ４がオフとされるタイミングt_1までは、音声区間とされて、その音声区間の音声データが音声認識の対象とされる。

【００２４】特徴抽出部５は、例えば、図３に示すように、パワースペクトラム分析部１１から構成され、フレーム化部２からの音声区間における観測ベクトルａとしての音声データをフーリエ変換することにより、音声の特徴量として、そのパワースペクトラムを求め、そのパワースペクトラムの各周波数成分をコンポーネントとする特徴ベクトルｙを算出する。特徴抽出部５で得られた特徴ベクトルｙは、音声認識部６に供給される。

【００２５】なお、パワースペクトラムの算出方法は、フーリエ変換によるものに限定されるものではない。すなわち、パワースペクトラムは、その他、例えば、いわゆるフィルタバンク法などによって求めることも可能である。

【００２６】また、ここでは、音声の特徴量として、パワースペクトラムを用いることとしているが、音声の特徴量としては、パワースペクトラムの他、ケプストラム係数（ＭＦＣＣ(Mel Frequency Cepstrum Coefficients)を含む）や、線形予測係数その他を採用することが可能である。

【００２７】音声認識部６は、特徴抽出部５から入力される特徴ベクトルｙを、所定数Ｋの音響モデルと１個の無音モデルの中のいずれかに分類し、その分類結果を、入力された音声の認識結果として出力する。すなわち、音声認識部６は、例えば、無音区間に対応する識別関数（特徴パラメータｙが無音モデルに分類されるかを識別するための関数）と、所定数Ｋの単語それぞれに対応する識別関数（特徴パラメータｙがいずれの音響モデルに分類されるかを識別するための関数）とを記憶しており、各音響モデルの識別関数の値を、特徴抽出部５からの特徴ベクトルｙを引数として計算する。そして、音声認識部６は、その関数値（いわゆるスコア）が最大である音響モデル（単語、または無音（ノイズ））を認識結果として出力する。

【００２８】即ち、図４は、図１の音声認識部６の詳細な構成例を示している。

【００２９】特徴抽出部５から入力される特徴ベクトルｙは、識別関数演算部２１－１乃至２１－ｋ、および識別関数演算部２１－ｓに供給される。識別関数演算部２１－ｋ（ｋ＝１，２，・・・，Ｋ）は、Ｋ個の音響モデルのうちのｋ番目に対応する単語を識別するための識別関数$G_k()$を記憶しており、特徴抽出部５からの特徴ベクトルｙを引数として、識別関数$G_k(y)$を演算する。識別関数演算部２１－ｓは、無音モデルに対応する無音区間を識別するための識別関数$G_s()$を記憶しており、特徴抽出部５からの特徴ベクトルｙを引数として、識別関数$G_s(y)$を演算する。

【００３０】なお、音声認識部６では、例えば、HMM(Hidden Markov Model)法を用いて、クラスとしての単語または無音の識別（認識）が行われる。

【００３１】ここで、図５は、HMMを示している。

【００３２】同図において、ＨＭＭは、Ｈ個の状態q_1乃至q_Hを有しており、状態の遷移は、自身への遷移と、右隣の状態への遷移のみが許されている。また、初期状態は、最も左の状態q_1とされ、最終状態は、最も右の状態q_Hとされており、最終状態q_Hからの状態遷移は禁止されている。このように、自身よりも左にある状態への遷移のないモデルは、left-to-rightモデルと呼ばれ、音声認識では、一般に、left-to-rightモデルが用いられる。

【００３３】いま、ＨＭＭのｋクラスを識別するためのモデル（音響モデル）を、ｋクラスモデルというとすると、ｋクラスモデルは、例えば、最初に状態q_kにいる確率（初期状態確率）$\pi_k(q_k)$、ある時刻（フレーム）ｔにおいて、状態q_iにいて、次の時刻ｔ＋１において、状態q_jに状態遷移する確率（遷移確率）$a_k(q_i, q_j)$、および状態q_jから状態遷移が生じるときに、その状態q_jが、特徴ベクトルＯを出力する確率＊（出力確率）$b_k(q_j)(O)$によって規定される（ｈ＝１，２，・・・，Ｈ）。

【００３４】そして、ある特徴ベクトル系列O_1，O_2，・・・が与えられた場合、例えば、そのような特徴ベクトル系列が観測される確率（観測確率）が最も高いモデルのクラスが、その特徴ベクトル系列の認識結果とされる。

【００３５】ここでは、この観測確率が、識別関数$G_k(y)$によって求められる。すなわち、識別関数$G_k(y)$は、特徴ベクトル（系列）$y=\{y_1, y_2, \cdots, y_T\}$に対する最適状態系列（最適な状態の遷移のしていき方）において、そのような特徴ベクトル（系列）$y=\{y_1, y_2, \cdots, y_T\}$が観測される確率を求めるものとして、次式（１）で与えられる。

【００３６】

【数１】

$$G_k(y)=\max_{q_1,q_2,\cdots,q_T}\pi_k(q_1)\cdot b_k(q_1)(y_1)\cdot a_k(q_1,q_2)\cdot b_k(q_2)(y_2)\cdots a_k(q_{T-1},q_T)\cdot b_k(q_T)(y_T)\quad\cdots(1)$$

【００３７】ここで、連続ＨＭＭにおいては、状態q_iにおける出力確率$b_i(q_i)(y_j)$は、確率分布によって表される。即ち、特徴ベクトル空間上のコンポーネントどうしに相関がないものとして、出力確率$b_i(q_i)(y_j)$を規定する確率分布に、正規分布関数※$P(q_i)(d)(y(t)(d))$を用いることとすると、その正規分布関数$P(q_i)(d)(y(t)(d))$は、次式で表すことができる。

【００３８】

【数２】

$$=\frac{1}{\sqrt{2\pi(\Sigma_k(q_i)(d,d)}}e^{-\frac{(\mu_k(q_i)(d)-y(t)(d))^2}{2(\Sigma_k(q_i)(d,d)}}\qquad k=1,2,\cdots,K:t=1,2,\cdots,T\quad\cdots(2)$$

但し、式（２）において、$\mu_k(q_i)(d)$は、正規分布を規定する平均ベクトルのｄ番目のコンポーネントを表し、$\Sigma_k(q_i)(d,d)$は、正規分布を規定する分散マトリクスの第ｄ行第ｄ列のコンポーネントを表す。また、$y(t)(d)$は、特徴ベクトル$y(t)$のｄ番目のコンポーネントを表す。

【００３９】ｋクラスモデルの状態q_iにおける出力確率は、式（２）の平均ベクトル$\mu_k(q_i)(d)$と、分散マトリクス$\Sigma_k(q_i)(d,d)$によって規定される。

【００４０】なお、ＨＭＭは、上述したように、初期状態確率$\pi_k(q_k)$、遷移確率$a_k(q_i, q_j)$、および出力確率$b_k(q_j)(O)$によって規定されるが、これらは、学習用の音声データから特徴ベクトルを算出し、その特徴ベクトルを用いて、予め求められる。

【００４１】また、ＨＭＭとして、図５に示したものを用いる場合には、常に、最も左の状態q_1から遷移が始まるので、状態q_1に対応する初期状態確率だけが１とされ、他の状態に対応する初期状態確率はすべて０とされる。

【００４２】さらに、ＨＭＭの学習方法としては、例えば、Baum-Welchの再推定法などが知られている。

【００４３】図４において、識別関数演算部２１－ｋ（ｋ＝１，２，・・・，Ｋ）は、ｋクラスモデルについて、あらかじめ学習により求められている初期状態確率$\pi_k(q_k)$、遷移確率$a_k(q_i, q_j)$、および出力確率$b_k(q_j)(O)$によって規定される式（２）の識別関数$G_k(y)$を記憶しており、特徴抽出部２からの特徴ベクトルｙを引数として、識別関数$G_k(y)$を演算し、その関数値（上述した観測確率）$G_k(y)$を、決定部２２に出力する。識別関数演算部２１－ｓは、音声モデルとしての、初期状態確率$\pi_s(q_k)$、遷移確率a

、（q_1，q_1）、および出力確率b_i（q_1）（O）によって規定される式（２）の識別関数G_i（y）と同様の識別関数G_i（y）を記憶しており、特徴抽出部２からの特徴ベクトルyを引数として、識別関数G_i（y）を演算し、その関数値（上述した観測確率）G_i（y）を、決定部２２に出力する。

【００４４】決定部２２では、識別関数演算部２１－１＊乃至２１－ｋ、および識別関数演算部２１－ｓそれぞれからの関数値G_i（y）（ここでは、関数値G_i（y）を含むものとする）に対して、例えば、次式（３）に示す決定規則を用いて、特徴ベクトルy、すなわち、入力された音声が属するクラス（音響モデル）が識別される。

$$C(y)=C_k,\ if\ G_k(y)=\max_i\{G_i(y)\} \quad \cdots(3)$$

【００４５】

【数３】

但し、C（y）は、特徴ベクトルyが属するクラスを識別する識別操作（処理）を行う関数を表す。また、式（３）の第２式の右辺におけるmaxは、それに続く関数値G_i（y）（ただし、ここでは、ｉ＝ｓ，１，２，・・・，Ｋ）の最大値を表す。

【００４６】決定部２２は、式（３）にしたがって、クラスを決定すると、対応する単語（または無音である旨）を、入力された音声の認識結果として出力する。

【００４７】図１に戻り、無音モデル補正部７は、ノイズ観測区間抽出部３から入力されるノイズ観測区間Ｔｎにおける音声データとしての環境ノイズに基づいて、音声認識部６に記憶されている無音モデルに対応する識別関数G（y）を生成し、この識別関数G（y）によって、音声認識部６に記憶されている無音モデルの適応を行う。

【００４８】具体的には、無音モデル補正部７は、ノイ※10※ズ観測区間抽出部３から入力されるノイズ観測区間Ｔｎの音声データ（環境ノイズ）のM個のフレームの各フレームについて、特徴ベクトルyの系列を観測し、その特徴ベクトルyの系列に対して統計的処理を施すことによって、無音モデルを規定する確率分布（無音モデルとしてのHMMの出力確率を規定する確率分布）を生成する。

【００４９】即ち、例えば、いま、無音モデルを規定する確率分布が正規分布で表されるとすると、無音モデル補正部７は、ノイズ観測区間Ｔｎのフレームの特徴ベクトルy（ｔ）の系列を用い、式（４）にしたがった計算を行うことにより、無音モデルG_i（y）としての正規分布を規定する平均値μ_{sil}と、分散マトリクスΣ_{sil}を求める。

【００５０】

【数４】

$$\mu_{sil}=\frac{1}{M}\sum_{t=1}^{M}y(t)$$

$$\Sigma_{sil}=\frac{1}{M}\sum_{t=1}^{M}(y(t)-\mu_{sil})(y(t)-\mu_{sil})^T \quad \cdots(4)$$

なお、式（４）におけるＴは、転置を表す。

【００５１】そして、無音モデル補正部７は、平均値μ_{sil}と分散マトリクスΣ_{sil}で規定される正規分布として の無音モデルG_i（y）によって、識別関数演算部２１－ｓの無音モデルG_i（y）としての識別関数を更新（補正）する。

【００５２】次に、図６のフローチャートを参照して、図１の音声認識装置による音声認識処理について説明する。

【００５３】フレーム化部２には、マイク１で集音された音声データが入力され、そこでは、音声データがフレーム化され、各フレームの音声データは、観測ベクトルａとして、ノイズ観測区間抽出部３、および特徴抽出部５に順次供給される。ノイズ観測区間抽出部３は、ステップＳ１において、フレーム化部２からの各フレームの音声データをバッファリングする。

【００５４】ここで、ノイズ観測区間抽出部３は、少なくともMフレーム以上の音声データを記憶することができる、図示せぬバッファを有しており、そのバッファの記憶容量分の音声データを記憶した後は、最も古い音声データに上書きする形で、新たな音声データを記憶するようになっている。従って、ノイズ観測区間抽出部３では、常に、最新のMフレーム以上の音声データが記憶される。

【００５５】その後、ステップＳ２において、音声区間が開始されたかどうか、即ち、ユーザによって、発話スイッチ４が操作されたかどうかが判定される。ステップＳ２において、音声区間が開始されていないと判定された場合、ステップＳ１に戻り、以下、同様の処理を繰り返す。

【００５６】また、ステップＳ２において、音声区間が開始されたと判定された場合、ステップＳ３に進み、無音モデル補正部７において、無音モデル適応処理が行われる。

【００５７】即ち、ステップＳ２では、ノイズ観測区間抽出部３は、発話スイッチ４がオンとされたタイミングｔ－の直前の区間であるノイズ観測区間Ｔｎの音声データ（環境ノイズ）を、その内蔵するバッファから抽出

し、無音モデル補正部７に供給する。

【００５８】無音モデル補正部７は、ノイズ観測区間Ｔｎの各フレームの音声データの特徴ベクトルｙ（ｔ）を求め、その特徴ベクトル（ｙ）を用いて、式（４）により、平均値μ_{wi}と分散マトリクスΣ_{wi}を求める。そして、無音モデル補正部７は、その平均値μ_{wi}と分散マトリクスΣ_{wi}で規定される正規分布Ｎ（μ_{wi}，Σ_{wi}）によって、音声認識部６の無音モデルＧｓ（ｙ）を更新する。

【００５９】一方、特徴抽出部５は、発話スイッチ４がオンとされ、音声区間が開始されると、フレーム化部２からの観測ベクトルａとしての音声データを音響分析し、その特徴ベクトルｙを求め、音声認識部６に供給する。音声認識部６は、ステップＳ４において、特徴抽出部５からの特徴ベクトルｙを用いて、無音と所定数Ｋの単語それぞれに対応する音響モデルの識別関数の値を演算し、ステップＳ５に進む。ステップＳ５では、音声認識部６は、ステップＳ５で演算した識別関数の関数値が最大となる音響モデルを選択し、対応する単語（または無音）を、音声の認識結果として出力する。

【００６０】その後、ステップＳ６に進み、音声認識処理を終了するかどうかが判定され、終了しないと判定された場合、ステップＳ１に戻り、次の発話について、以下、同様の処理が行われる。

【００６１】また、ステップＳ６において、音声認識処理を終了すると判定された場合、即ち、例えば、ユーザが、音声認識装置の電源をオフする操作を行った場合、処理を終了する。

【００６２】次に、上述の場合においては、図７に示すように、各発話の音声区間ごとに、その直前のノイズ観測区間Ｔｎの音声データ（環境ノイズ）のみに基づいて、無音モデルの適応が行われる。即ち、いま、音声認識装置において音声認識処理が開始されてから、ユーザが行った発話を、第１発話、第２発話、・・・とカウントすることとし、第Ｎ発話を、現在の音声認識の対象となっている音声の音声区間（注目音声区間）の発話であるとすると、第Ｎ発話の音声区間である注目音声区間の音声の認識には、その注目音声区間の直前のノイズ観測区間Ｔｎの環境ノイズだけに基づいて生成された無音モデルが用いられる。

【００６３】ここで、図７において（後述する図８，図１０，図１２においても同様）、Ｉｎは、第ｎ発話を表し、Ｇｓ［Ｉｎ］は、第ｎ発話の音声区間の音声の認識に用いられる無音モデルを表す。

【００６４】注目音声区間の直前のノイズ観測区間Ｔｎの環境ノイズだけに基づいて、無音モデルを生成する場合、前述したように、例えば、注目音声区間の直前において、ユーザが、音声の入力に用いるマイクを叩く等したときや、観衆が拍手を行ったとき等の、いわば突発的なノイズが生じたときには、その突発的なノイズに基づいて、無音モデルの適応が行われる。

【００６５】しかしながら、注目音声区間の、ユーザが発話を行う区間においては、突発的なノイズが存在しなくなるから、突発的なノイズに基づいて生成された無音モデルを用いて、注目音声区間の音声を認識したのでは、認識率が劣化することがある。

【００６６】また、比較的定常的なノイズが長時間連続している場合には、注目音声区間のみならず、過去に音声認識の対象とされた音声の音声区間の直前の区間における環境ノイズにも基づいて、無音モデルを生成した方が、環境ノイズをより的確に表す無音モデルを得ることができると予想され、さらに、そのような無音モデル用いて、注目音声区間の音声認識を行うことにより、精度の高い音声認識を行うことが可能となる。

【００６７】そこで、音声認識装置では、注目音声区間の直前の環境ノイズだけでなく、過去の１以上の音声区間の直前の環境ノイズにも基づいて、以下のような第１乃至第３の３つの適応方法のいずれかにより、注目音声区間の音声を認識するのに用いる無音モデルの適応を行うことが可能となっている。

【００６８】即ち、第１の適応方法では、図８に示すように、第Ｎ発話の音声区間である注目音声区間の直前の環境ノイズと、過去の第１乃至第Ｎ－１発話の音声区間の直前の環境ノイズから、注目音声区間の音声の認識に用いる無音モデルＧｓ［Ｉｎ］が生成される。

【００６９】この場合、図６のステップＳ３における無音モデル適応処理は、図９のフローチャートに示すように行われる。

【００７０】即ち、この場合、ステップＳ１１において、無音モデル補正部７は、第１乃至第Ｎ発話の音声区間の直前のノイズ観測区間Ｔｎの音声データ（環境ノイズ）の特徴ベクトルｙ（ｔ）を計算する。従って、この場合、ノイズ観測区間抽出部３では、注目音声区間である第Ｎ発話の音声区間の直前の環境ノイズだけでなく、過去の第１乃至第Ｎ発話の音声区間の直前の環境ノイズも記憶しておく必要がある。

【００７１】さらに、無音モデル補正部７は、第１乃至第Ｎ発話の音声区間の直前のノイズ観測区間Ｔｎの音声データ（環境ノイズ）の特徴ベクトルｙ（ｔ）の集合の平均ベクトルμ_{wi}と分散マトリクスΣ_{wi}を、次式にしたがって計算し、その平均ベクトルμ_{wi}と分散マトリクスΣ_{wi}によって規定される正規分布を、注目音声区間の音声の認識に用いる無音モデルＧｓ［Ｉｎ］とする。

【００７２】

【数５】

13　14

$$\mu_{sil}=\frac{1}{\sum_{i=1}^{N}M(i)}\left[\sum_{i=1}^{N}W_i\cdot\sum_{t=1}^{M(i)}y(t)\,[I_i]\right]$$

$$\Sigma_{sil}=\frac{1}{\sum_{i=1}^{N}M(i)}\left\{\sum_{i=1}^{N}\cdot\sum_{t=1}^{M(i)}\left[W_i\cdot y(t)\,[I_i]-\mu_{sil}\right]\left[W_i\cdot y(t)\,[I_i]-\mu_{sil}\right]^T\right\}$$

・・・（５）

【００７３】なお、Ｍ（ｉ）は、第ｉ発話の音声区間の直前のノイズ観測区間Ｔｎのフレーム数を表し、本実施の形態では、上述したことから、すべてＭフレームである。但し、ノイズ観測区間Ｔｎのフレーム数は、各発話ごとに、異なるフレーム数とすることが可能である。

【００７４】また、w_iは、第ｉ発話の音声区間の直前の環境ノイズに対する重みを表す。この重みw_iは、式

$$\sum_{i=1}^{N}W_i=1$$

＊（６）を満たすもので、例えば、第Ｎ発話の音声区間（注目音声区間）の直前の環境ノイズに対する重みw_Nは、０．５とし、第１乃至第Ｎ－１発話の音声区間の直前の環境ノイズに対する重みw_1乃至w_{N-1}は、いずれも、０．５／（Ｎ－１）とすることが可能である。

【００７５】

【数６】

・・・（６）

【００７６】さらに、重みw_iは、注目音声区間である第Ｎ発話の音声区間から離れた音声区間の直前の環境ノイズに対するものほど、小さな値にするようにすること等が可能である。

【００７７】また、式（５）において、ｙ（ｔ）［Ｉｉ］は、第ｉ発話の音声区間の直前の環境ノイズの第ｔフレーム（ノイズ観測区間Ｔｎの第ｔフレーム）の特徴ベクトルを表す。

【００７８】次に、第２の適応方法では、図１０に示すように、過去の第１乃至第Ｎ－１発話の音声区間の直前の環境ノイズに基づいて、第１の無音モデルG_{w1}が生成されるとともに、第Ｎ発話の音声区間である注目音声区間の直前の環境ノイズに基づいて、第２の無音モデルG_{w2}が生成され、その第１の無音モデルG_{w1}と、第２の無音モデルG_{w2}とに基づいて、注目音声区間の音声の認識に用いる無音モデルG_{α}［I_{α}］が生成される。

【００７９】この場合、図６のステップＳ３における無音モデル適応処理は、図１１のフローチャートに示すように行われる。

【００８０】即ち、この場合、ステップＳ２１において、無音モデル補正部７は、第１乃至第Ｎ－１発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズの特徴ベクトルｙ（ｔ）を計算する。さらに、無音モデル補正部７は、第１乃至第Ｎ－１発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズの特徴ベクトルｙ（ｔ）の集合の平均ベクトルμ_{sil-1}と分散マトリクスΣ_{sil-1}を、式（５）における場合と同様に計算し、その平均ベクトルμ_{sil-1}と分散マトリクスΣ_{sil-1}によって規定される正規分布を、第１の無音モデルG_{w1}とする。

＊

$$\mu_{sil}=a_{\mu1}\mu_{sil-1}+b_{\mu2}\mu_{sil-2}$$

$$\Sigma_{sil}=a_{\Sigma1}\Sigma_{sil-1}+b_{\Sigma2}\Sigma_{sil-2}$$

【００８５】ここで、式（７）における$a_{\mu i}$，$b_{\mu i}$，$a_{\Sigma i}$，$b_{\Sigma i}$は、いずれも、０以上１以下の範囲の値を

＊【００８１】そして、ステップＳ２２に進み、無音モデル補正部７は、注目フレームである第Ｎ発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズの特徴ベクトルｙ（ｔ）を計算する。さらに、無音モデル補正部７は、第Ｎ発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズの特徴ベクトルｙ（ｔ）の集合の平均ベクトルμ_{sil-2}と分散マトリクスΣ_{sil-2}を、上述の式（４）にしたがって計算し、その平均ベクトルμ_{sil-2}と分散マトリクスΣ_{sil-2}によって規定される正規分布を、第２の無音モデルG_{w2}とする。

【００８２】以上のようにして、第１の無音モデルG_{w1}と、第２の無音モデルG_{w2}を得た後は、ステップＳ２３に進み、無音モデル補正部７は、第１の無音モデルG_{w1}と、第２の無音モデルG_{w2}とを統合することにより、注目音声区間の音声の認識に用いる無音モデルG_{α}［I_{α}］を生成する。

【００８３】即ち、無音モデル補正部７は、例えば、式（７）にしたがい、第１の無音モデルG_{w1}を規定する平均ベクトルμ_{sil-1}と、第２の無音モデルG_{w2}を規定する平均ベクトルμ_{sil-2}とを統合し、平均ベクトルμ_{sil}を求めるとともに、第１の無音モデルG_{w1}を規定する分散マトリクスΣ_{sil-1}と、第２の無音モデルG_{w2}を規定する分散マトリクスΣ_{sil-2}とを統合し、分散マトリクスΣ_{sil}を求める。そして、無音モデル補正部７は、その平均ベクトルμ_{sil}と分散マトリクスΣ_{sil}によって規定される正規分布を、注目音声区間の音声の認識に用いる無音モデルG_{α}［I_{α}］とする。

【００８４】

【数７】

・・・（７）

とる重みであり、式$a_{\beta 1}+b_{\beta 2}=1$と、式$a_{\xi 1}+b_{\xi 2}=1$を満たすものである。

【００８６】環境ノイズが、比較的定常的なものである場合には、重み$a_{\beta 1}$，$b_{\beta 2}$，$a_{\xi 1}$，$b_{\xi 2}$としては、例えば、同一の値を使用することができる。また、環境ノイズが、時間の経過に伴って、比較的変化する場合には、重み$a_{\beta 1}$，$b_{\beta 2}$，$a_{\xi 1}$，$b_{\xi 2}$としては、例えば、$a_{\beta 1}$と$a_{\xi 1}$については、小さな値を、$b_{\beta 2}$と$b_{\xi 2}$については、大きな値を、それぞれ採用することができる。さらに、注目音声区間の直前の環境ノイズが、突発的なものである場合には、重み$a_{\beta 1}$，$b_{\beta 2}$，$a_{\xi 1}$，$b_{\xi 2}$としては、例えば、$a_{\beta 1}$と$a_{\xi 1}$については、大きな値を、$b_{\beta 2}$と$b_{\xi 2}$については、小さな値を、それぞれ採用することができる。

【００８７】なお、第１および第２の適応方法においては、過去の音声区間については、注目音声区間より過去の音声区間すべての直前の環境ノイズを用いる他、そのうちの一部の音声区間の直前の環境ノイズを用いて、注目音声区間の音声認識に用いる無音モデルの適応を行うようにすることが可能である。

【００８８】次に、第３の適応方法では、図１２に示すように、第Ｎ発話の音声区間である注目音声区間の直前の環境ノイズに基づいて、無音モデルが生成され、その無音モデルと、過去の音声区間、即ち、図１２の実施の形態では、注目音声区間の直前の音声区間（第Ｎ－１発話の音声区間）の音声認識に用いられた無音モデルとに基づいて、注目音声区間の音声の認識に用いる無音モデルＧ，［I_N］が生成される。

【００８９】この場合、図６のステップＳ３における無音モデル適応処理は、図１３のフローチャートに示すように行われる。

【００９０】即ち、この場合、ステップＳ３１において、無音モデル補正部７は、直前の発話、つまり第Ｎ－１発話の音声区間の音声認識に用いられた無音モデルＧ，［I_{N-1}］を、音声認識部６（図４）から取得し、ステップＳ３２に進む。

【００９１】ステップＳ３２では、無音モデル補正部７は、注目フレームである第Ｎ発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズの特徴ベクトルｙ（ｔ）を計算する。さらに、無音モデル補正部７は、第Ｎ発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズの特徴ベクトルｙ（ｔ）の集合の平均ベクトルと分散マトリクスを、上述の式（４）にしたがって計算し、その平均ベクトルと分散マトリクスによって規定される正規分布としての無音モデルＧ，’［I_V］を生成する。

【００９２】そして、ステップＳ３３に進み、無音モデル補正部７は、第Ｎ－１発話の音声区間の音声認識に用いられた無音モデルＧ，［$I_{V'}$］と、第Ｎ発話の音声区間の直前の環境ノイズだけから得られた無音モデルＧ，’［I_N］とを統合することにより、注目音声区間の音声の認識に用いる無音モデルＧ，［I_V］を生成する。

【００９３】即ち、例えば、第Ｎ－１発話の音声区間の音声認識に用いられた無音モデルＧ，［$I_{V'}$］としての正規分布を規定する平均ベクトルと分散マトリクスを、それぞれμ_{vst-1}　とΣ_{vst-1}とするとともに、第Ｎ発話の音声区間の直前の環境ノイズだけから得られた無音モデルＧ，’［I_N］としての正規分布を規定する平均ベクトルと分散マトリクスを、それぞれμ_{vst-2}　とΣ_{vst-2}とすると、無音モデル補正部７は、ステップＳ３３において、例えば、上述の式（７）にしたがい、平均ベクトルμ_{vst-1}　とμ_{vst-2}　とを統合し、平均ベクトルμ_{vst}を求めるとともに、分散マトリクスΣ_{vst-1}とΣ_{vst-2}とを統合し、分散マトリクスΣ_{vst}を求める。そして、無音モデル補正部７は、その平均ベクトルμ_{vst}と分散マトリクスΣ_{vst}によって規定される正規分布を、注目音声区間の音声の認識に用いる無音モデルＧ，［I_V］とする。

【００９４】なお、第３の適応方法においては、第Ｎ－１発話の音声認識に用いられた無音モデルの他、過去の他の発話の音声認識に用いられた無音モデルを用いて、注目音声区間の音声認識に用いる無音モデルを生成することが可能である。

【００９５】以上のように、注目音声区間の直前の環境ノイズだけでなく、過去の１以上の音声区間の直前の環境ノイズにも基づいて、注目音声区間の音声を認識するのに用いる無音モデルの適応を行うようにしたので、音声区間中の無音部分に十分対処可能な無音モデルの適応を行うことができ、これにより、無音部分に起因する音声認識性能の劣化を防止（低減）することができる。

【００９６】ところで、ノイズ環境下において音声を認識する場合の特徴量（特徴ベクトル）の抽出方法の１つに、例えば、スペクトルサブトラクション（Spectral Subtraction）と呼ばれるものがある。

【００９７】スペクトルサブトラクションでは、音声の発話がされる前の入力（音声区間の前の入力）を、ノイズとして、そのノイズの平均スペクトルが算出される。そして、音声区間の音声から、ノイズの平均スペクトルが差し引かれ（Subtract）、その残りを、真の音声成分として、特徴ベクトルが算出される。

【００９８】一方、図１の音声認識装置における特徴抽出部５では、各フレームの音声データとしての観測ベクトルａから、特徴ベクトルが求められるが、このことは、観測ベクトル空間上の、ある点を表す観測ベクトルａを、特徴ベクトル空間上に写像することにより、その特徴ベクトル空間上の、対応する点を表す特徴ベクトルに変換する処理が行われると考えることができる。

【００９９】従って、特徴ベクトルは、特徴ベクトル空間上の、ある１点（観測ベクトルａに対応する点）を表す。

【０１００】スペクトルサブトラクションでは、観測ベクトルａから、ノイズの平均スペクトル成分が取り除か

れて、特徴ベクトルが算出されるが、この特徴ベクトルは、上述したように、特徴ベクトル空間上の１点であるため、ノイズの平均的な性質を考慮したものとはなっているが、ノイズの分散などの不規則な性質を考慮したものとはなっていない。

【０１０１】このため、スペクトルサブトラクション処理後に得られる特徴ベクトルは、観測ベクトルａの特徴を充分に（あるいは、正確に）表現しているとはいえず、そのような特徴ベクトルでは、認識性能を十分に向上させることができないことがある。

【０１０２】そこで、図１４は、本発明を適用した音声認識装置の他の一実施の形態の構成例を示している。なお、図中、図１における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。

【０１０３】即ち、図１４の実施の形態では、図１の特徴抽出部５、音声認識部６、または無音モデル補正部７に替えて、特徴抽出部３１、音声認識部３２、または無音モデル補正部３３がそれぞれ設けられており、さらに、ノイズ観測区間抽出部３が出力する環境ノイズが、無音モデル補正部３３だけでなく、特徴抽出部３１にも供給されるようになっている。

【０１０４】但し、ノイズ観測区間抽出部３は、フレーム化部２から入力されるフレーム単位の音声データを、図１における場合よりも長い時間（例えば、２Ｍフレーム分以上など）だけバッファリングすることができるようになっている。

【０１０５】即ち、図１４の実施の形態においては、ノイズ観測区間抽出部３は、例えば、図１５に示すように、発話スイッチ４がオンとされたタイミングt_0からＭフレーム分だけ以前のタイミングt_sまでを、ノイズ観測区間Ｔｎとするとともに、さらの、そのノイズ観測区間ＴｎからＭフレーム分だけ以前のタイミングt_0までをノイズ観測区間Ｔｍとして、その連続する２つのノイズ観測区間ＴｎとＴｍにおける２Ｍフレーム分の観測ベクトルａを抽出して、特徴抽出部３１、および無音モデル補正部３３に出力する。

【０１０６】なお、２つのノイズ観測区間ＴｎとＴｍは、連続していなくてもかまわない。また、ノイズ観測区間Ｔｎは、上述したように、無音モデルの適応を行うための環境ノイズを得るための区間であるが、ノイズ観測区間Ｔｍは、後述する特徴分布を抽出するための環境ノイズを得るための区間である。さらに、ここでは、２＊

$$x(t) = y(t) - u(t)$$

【０１１３】ただし、ここでは、環境ノイズが不規則な特性を有し、また、観測ベクトルａ（ｔ）としての音声データは、真の音声成分に環境ノイズを加算したものであると仮定している。

【０１１４】一方、ノイズ観測区間抽出部３から入力される音声データとしてのノイズ観測区間Ｔｍにおける環境ノイズは、特徴抽出部３１において、ノイズ特性算出部４３に供給される。ノイズ特性算出部４３では、ノイズ観測区間Ｔｍにおける環境ノイズの特性が求められる。

＊つのノイズ観測区間ＴｍとＴｎを、いずれも、Ｍフレームで構成するようにしたが、ノイズ観測区間ＴｍとＴｎのフレーム数は、同一である必要はない。

【０１０７】特徴抽出部３１は、ノイズ観測区間抽出部３から入力されるノイズ観測区間ＴｍとＴｎのうちの前半のノイズ観測区間Ｔｍの環境ノイズだけが存在する音声データに基づいて、フレーム化部２から入力される、タイミングt_s以降の音声区間の観測ベクトルａから環境ノイズ成分を除去して、その特徴量を抽出する。

【０１０８】即ち、特徴抽出部３１は、例えば、図１の特徴抽出部５と同様に、観測ベクトルａとしての音声データをフーリエ変換し、そのパワースペクトラムを求め、そのパワースペクトラムの各周波数成分をコンポーネントとする特徴ベクトルｙを算出する。さらに、特徴抽出部３１は、観測ベクトルａとしての音声データに含まれる真の音声成分を、その特徴量の空間（特徴ベクトル空間）に写像したときに得られる、その特徴ベクトル空間上の分布を表すパラメータ（以下、特徴分布パラメータと記述する）Ｚを、特徴ベクトルｙとノイズ観測区間Ｔｍの環境ノイズに基づいて算出し、音声認識部３２に供給する。

【０１０９】即ち、図１６は、図１４の特徴抽出部３１の詳細な構成例を示している。なお、図中、図３の特徴抽出部５における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、特徴抽出部３１は、特徴分布パラメータ算出部４２とノイズ特性算出部４３が新たに設けられている他は、図３の特徴抽出部５と同様に構成されている。

【０１１０】フレーム化部２から入力される観測ベクトルａは、特徴抽出部３１において、パワースペクトラム分析部１１に供給され、特徴ベクトルｙとしてのパワースペクトラムとされる。なお、ここでは、１フレームの音声データとしての観測ベクトルａが、Ｄ個のコンポーネントからなる特徴ベクトル（Ｄ次元の特徴ベクトル）に変換されるものとする。

【０１１１】ここで、第ｔフレームの観測ベクトルａ（ｔ）から得られる特徴ベクトルｙ（ｔ）のうち、真の音声のスペクトル成分をｘ（ｔ）と、環境ノイズのスペクトル成分をｕ（ｔ）と表す。この場合、真の音声のスペクトル成分ｘ（ｔ）は、次式（８）で表される。

【０１１２】

【数８】

$$\cdots (8)$$

【０１１５】即ち、ここでは、音声区間における環境ノイズのパワースペクトラムｕ（ｔ）の分布が、その音声

区間の直前のノイズ観測区間Ｔｍにおける環境ノイズと同一であり、かつ、その分布が正規分布であると仮定して、ノイズ特性算出部４３において、その正規分布を規定する、環境ノイズの平均ベクトルμ’と分散マトリクス＊スΣ’が、次式（９）にしたがって求められる。

【０１１６】

【数９】

$$\mu'(i) = \frac{1}{M}\sum_{t=1}^{M} y(t)(i)$$

$$\Sigma'(i,j) = \frac{1}{M}\sum_{t=1}^{M} (y(t)(i)-\mu'(i))(y(t)(j)-\mu'(j)) \quad \cdots (9)$$

【０１１７】ただし、μ’（ｉ）は、平均ベクトルμ’のｉ番目のコンポーネントを表す（ｉ＝１，２，・・・，Ｄ）。また、ｙ（ｔ）（ｉ）は、第ｔフレームの特徴ベクトルのｉ番目のコンポーネントを表す。さらに、Σ’（ｉ，ｊ）は、分散マトリクスΣ’の、第ｉ行、第ｊ列のコンポーネントを表す（ｊ＝１，２，・・・，Ｄ）。

※【０１１８】ここで、計算量の低減のために、環境ノイズについては、特徴ベクトルｙの各コンポーネントが、互いに無相関であると仮定する。この場合、次式に示すように、分散マトリクスΣ’は、対角成分以外は０となる。

【０１１９】

【数１０】

$$\Sigma'(i,j) = 0,\ i \neq j \quad \cdots (10)$$

【０１２０】なお、環境ノイズについて、特徴ベクトルｙの各コンポーネントが、互いに無相関であると仮定しなくても、計算量は増加するが、以下説明する処理を行うことは可能である。

【０１２１】ノイズ特性算出部４３は、以上のようにして、環境ノイズの特性としての、正規分布を規定する平均ベクトルμ’および分散マトリクスΣ’を求め、特徴分布パラメータ算出部４２に供給する。

【０１２２】一方、パワースペクトラム分析部１１の出力、すなわち、環境ノイズを含む音声区間の音声の特徴ベクトルｙも、特徴分布パラメータ算出部４２に供給される。特徴分布パラメータ算出部４２は、パワースペクトラム分析部１１からの特徴ベクトルｙ、およびノイズ★

★特性算出部４３からの環境ノイズの特性（ここでは、環境ノイズを表す正規分布を規定する平均ベクトルμ’と分散マトリクスΣ’）に基づいて、真の音声のパワースペクトラムの分布（推定値の分布）を表す特徴分布パラメータを算出する。

【０１２３】即ち、特徴分布パラメータ算出部４２は、真の音声のパワースペクトラムの分布が正規分布であるとして、その平均ベクトルξと分散マトリクスΨを、特徴分布パラメータとして、次式（１１）乃至（１４）にしたがって計算する。

【０１２４】

【数１１】

$$\begin{aligned}
\xi(t)(i) &= E[y(t)(i)-u(t)(i)] \\
&= \int_0^{y(t)(i)} (y(t)(i)-u(t)(i)) \frac{P(u(t)(j))}{\int_0^{y(t)(i)} P(u(t)(i))\,du(t)(i)}\,du(t)(i) \\
&= \frac{y(t)(i)\int_0^{y(t)(i)} P(u(t)(i))\,du(t)(i) - \int_0^{y(t)(i)} u(t)(i)P(u(t)(i))\,du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i))\,du(t)(i)} \\
&= y(t)(i) - \frac{\int_0^{y(t)(i)} u(t)(i)P(u(t)(i))\,du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i))\,du(t)(i)} \quad \cdots (11)
\end{aligned}$$

【０１２５】

【数１２】

21　　22

i＝jのとき

$$\Psi(t)(i,j) = V[x(t)(i)]$$
$$= E[(x(t)(i))^2] - (E[x(t)(i)])^2$$
$$(= E[(x(t)(i))^2] - (\xi(t)(i))^2)$$

i≠jのとき

$$\Psi(t)(i,j) = 0 \qquad \cdots (12)$$

【０１２６】　＊10＊【数１３】

$$E[(x(t)(i))^2] = E[(y(t)(i) - u(t)(i))^2]$$

$$= \int_0^{y(t)(i)} (y(t)(i) - u(t)(i))^2 \frac{P(u(t)(j))}{\int_0^{y(t)(i)} P(u(t)(i))du(t)(i)} du(t)(i)$$

$$= \frac{1}{\int_0^{y(t)(i)} P(u(t)(i))du(t)(i)} \times \left\{ (y(t)(i))^2 \int_0^{y(t)(i)} P(u(t)(i))du(t)(i) \right.$$

$$-2y(t)(i) \int_0^{y(t)(i)} u(t)(i)P(u(t)(i))du(t)(i)$$

$$\left. + \int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i))du(t)(i) \right\}$$

$$= (y(t)(i))^2 - 2y(t)(i) \frac{\int_0^{y(t)(i)} u(t)(i)P(u(t)(i))du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i))du(t)(i)}$$

$$+ \frac{\int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i))du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i))du(t)(i)} \qquad \cdots (13)$$

【０１２７】　※　※【数１４】

$$P(u(t)(i)) = \frac{1}{\sqrt{2\pi\,\Sigma'(i,i)}} e^{-\frac{1}{2\Sigma'(i,i)}(u(t)(i) - \mu'(i))^2} \qquad \cdots (14)$$

【０１２８】ここで、ξ（ｔ）（ｉ）は、第ｔフレームにおける平均ベクトルξ（ｔ）のｉ番目のコンポーネントを表す。また、Ｅ［］は、［］内の平均値を意味する。ｘ（ｔ）（ｉ）は、第ｔフレームにおける真の音声のパワースペクトラムｘ（ｔ）のｉ番目のコンポーネントを表す。さらに、ｕ（ｔ）（ｉ）は、第ｔフレームにおける環境ノイズのパワースペクトラムのｉ番目のコンポーネントを表し、Ｐ（ｕ（ｔ）（ｉ））は、第ｔフレームにおける環境ノイズのパワースペクトラムのｉ番目のコンポーネントがｕ（ｔ）（ｉ）である確率を表す。ここでは、環境ノイズの分布として正規分布を仮定しているので、Ｐ（ｕ（ｔ）（ｉ））は、式（１４）に示したように表される。

【０１２９】また、Ψ（ｔ）（ｉ，ｊ）は、第ｔフレームにおける分散マトリクスΨ（ｔ）の、第ｉ行、第ｊ列のコンポーネントを表す。さらに、Ｖ［］は、［］内の分散を表す。

【０１３０】特徴分布パラメータ算出部４２は、以上のようにして、各フレームごとに、平均ベクトルξおよび分散マトリクスΨを、真の音声の特徴ベクトル空間上での分布（ここでは、真の音声の特徴ベクトル空間上での分布が正規分布であると仮定した場合の、その分布）を表す特徴分布パラメータとして求める。

【０１３１】特徴分布パラメータ算出部４２は、音声区間の各フレームについて求めた特徴分布パラメータを、音声認識部３２に出力する。すなわち、いま、音声区間

がＴフレームであったとし、そのＴフレームそれぞれにおいて求められた特徴分布パラメータを、ｚ（ｔ）＝｛ξ（ｔ），Ψ（ｔ）｝（ｔ＝１，２，・・・，Ｔ）と表すと、特徴分布パラメータ算出部４２は、特徴分布パラメータ（系列）Ｚ＝｛ｚ（１），ｚ（２），・・・，ｚ（Ｔ）｝を、音声認識部３２に供給する。

【０１３２】図１４に戻り、音声認識部３２は、特徴抽出部３１から入力される特徴分布パラメータＺを、所定数Ｋの音響モデルと１個の無音モデルのうちのいずれかに分類し、その分類結果を、入力された音声の認識結果として出力する。

【０１３３】即ち、音声認識部３２は、例えば、無音区間に対応する識別関数と、所定数Ｋの単語それぞれに対応する識別関数とを記憶しており、各音響モデルの識別関数の値を、特徴抽出部３１からの特徴分布パラメータＺを引数として計算する。そして、その関数値が最大である音響モデル（単語、または無音（ノイズ））が認識結果として出力される。

【０１３４】ここで、図１７は、図１４の音声認識部３２の詳細な構成例を示している。なお、図中、図４の音声認識部６における場合と対応する部分については、同一の符号を付してある。即ち、音声認識部３２は、基本＊＊的に、図４の音声認識部６と同様に構成されている。

【０１３５】但し、識別関数演算部２１－１乃至２１－ｋ、および識別関数演算部２１－ｓには、特徴抽出部３１の特徴分布パラメータ算出部４２が出力する特徴分布パラメータＺが供給されるようになっており、識別関数演算部２１－ｋ（ｋ＝１，２，・・・，Ｋ，ｓ）は、特徴分布パラメータＺを引数とする識別関数Ｇｋ（Ｚ）を、音響モデルとして記憶している。

【０１３６】図１７の実施の形態において、音声認識部３２が、例えば、図４の音声認識部６と同様に、ＨＭＭ法を用いて、クラスとしての単語または無音の識別（認識）を行う場合、音声認識部３２は、音響モデルとしてのＨＭＭにおいて、特徴分布パラメータの系列Ｚ＝｛ｚ１，ｚ２，・・・，ｚＴ｝が観測される観測確率を、識別関数Ｇｋ（Ｚ）によって求める。即ち、この場合、識別関数Ｇｋ（Ｚ）は、特徴分布パラメータの系列Ｚ＝｛ｚ１，ｚ２，・・・，ｚＴ｝に対する最適状態系列において、そのような特徴分布パラメータ（の系列）Ｚ＝｛ｚ１，ｚ２，・・・，ｚＴ｝が観測される確率を求めるものとして、次式（１５）で与えられる。

【０１３７】

【数１５】

$$G_k(Z) = \max_{q_1, q_2, \cdots, q_T} \pi_k(q_1) \cdot b_k'(q_1)(z_1) \cdot a_k(q_1, q_2) \cdot b_k'(q_2)(z_2) \cdots a_k(q_{T-1}, q_T) \cdot b_k'(q_T)(z_T) \quad \cdots (15)$$

【０１３８】ここで、$b_i'(q_i)(z_j)$は、出力がz_jで表される分布であるときの出力確率を表す。式（１）で説明したように、状態遷移時に各特徴ベクトルを出力する確率である出力確率$b_i(S)(O_t)$に（Ｓは状態を表す）、特徴ベクトル空間上のコンポーネントに相関がないものとして、正規分布関数を用いることとした場合、入力がz_cで表される分布であるときは、出力確率$b_i'(S)(z_c)$は、平均ベクトル$\mu_i(S)$※※と分散マトリクス$\Sigma_i(S)$とによって規定される確率密度関数$P_{is}(S)(x)$、および第ｔフレームの特徴ベクトル（ここでは、パワースペクトラム）ｘの分布を表す確率密度関数$P_f(t)(x)$を用いて、次式（１６）により求めることができる。

【０１３９】

【数１６】

$$b_k'(S)(z_t) = \int P^f(t)(x) P_k^m(S)(x)\, dx = \prod_{i=1}^{D} P(S)(i)(\xi(t)(i), \Psi(t)(i,i)) \quad k=1,2,\cdots,K: S=q_1,q_2\cdots,q_T: t=1,2\cdots,T \quad \cdots (16)$$

【０１４０】ただし、式（１６）における積分の積分区間は、Ｄ次元の特徴ベクトル空間（ここでは、パワースペクトラム空間）の全体である。

【０１４１】また、式（１６）において、Ｐ（Ｓ）（ｉ）（ξ（ｔ）（ｉ），Ψ（ｔ）（ｉ，ｉ））は、次式（１７）で表される。

【０１４２】

【数１７】

25　　26

$$P(S)(i)(\xi(t)(i),\Psi(t)(i,i)) = \frac{1}{\sqrt{2\pi(\Sigma_k(S)(i,i)+\Psi(t)(i,i))}} e^{-\frac{(\mu_k(S)(i)-\xi(t)(i))^2}{2(\Sigma_k(S)(i,i)+\Psi(t)(i,i))}} \quad \cdots (17)$$

【０１４３】ただし、$\mu_k(S)(i)$は、平均ベクトル$\mu_k(S)$のｉ番目のコンポーネントを、$\Sigma_k(S)(i,i)$は、分散マトリクス$\Sigma_k(S)$の、第ｉ行第ｉ列のコンポーネントを、それぞれ表す。そして、ｋクラスモデルの出力確率は、これらによって規定される。

【０１４４】なお、ＨＭＭは、上述した場合と同様に、学習用の音声データから特徴ベクトルを算出し、その特徴ベクトルを用いて、予め求めておく。

【０１４５】ここで、特徴分布パラメータＺに基づく音声認識に用いられる出力確率を規定する式（１７）の確率分布は、特徴分布パラメータＺの分散Ψ（ｔ）（ｉ，ｉ）をＯとすると、特徴ベクトルの分散を考慮しない場合の連続ＨＭＭにおける出力確率を規定する式（２）の確率分布に一致する。

【０１４６】決定部２２は、図４における場合と同様に、識別関数演算部２１－１乃至２１－ｋ、および識別関数演算部２１－ｓそれぞれからの関数値$G_k(Z)$（関数値$G_k(Z)$を含む）に対して、上述の式（３）と同様の決定規則を用いて、特徴分布パラメータＺ、即ち、入力された音声が属するクラス（音響モデル）を識＊別し、音声認識結果として出力する。

【０１４７】図１４に戻り、無音モデル補正部３３は、ノイズ観測区間抽出部３から入力されるノイズ観測区間ＴｍとＴｎにおける音声データとしての環境ノイズに基づいて、音声認識部３２に記憶されている無音モデルに対応する識別関数Ｇ$_k$（Ｚ）を生成し、この識別関数Ｇ$_k$（Ｚ）によって、音声認識部３２に記憶されている無音モデルの適応を行う。

【０１４８】具体的には、無音モデル補正部３３では、ノイズ観測区間抽出部３から入力される後半のノイズ観測区間Ｔｎの音声データ（環境ノイズ）のＭ個のフレームの各フレームについて、特徴ベクトルｙが観測され、さらに、特徴抽出部３１における場合と同様にして、前半のノイズ観測区間Ｔｍにおける環境ノイズを用いて、後半のノイズ観測区間Ｔｎの各フレーム＃ｉにおける環境ノイズの特徴分布パラメータの、次式で示される系列が生成される。

【０１４９】

【数１８】

$$\{F_1(y), F_2(y), \cdots, F_M(y)\} \quad \cdots (18)$$

【０１５０】ここで、環境ノイズの特徴分布パラメータ$F_i(y)$は、ユーザの音声のない部分、つまり無音（正確には、環境ノイズが存在する）の特徴ベクトルの分布を表すから、以下、適宜、無音特徴分布とも記述する。

※【０１５１】次に、無音モデル補正部３３は、無音特徴分布を、次式（１９）に従い、無音モデルに対応する確率分布Ｇ$_k$（Ｚ）に写像する。

【０１５２】

【数１９】

$$※ \quad G_s(y) = V(F_1(y), F_2(y), \cdots, F_M(y)) \quad \cdots (19)$$

【０１５３】但し、Ｖは無音特徴分布｛F_i（Ｚ），ｉ＝１，２，・・・，Ｍ｝を無音モデルＧ$_k$（Ｚ）に写像する補正関数（写像関数）である。

【０１５４】この写像は、無音特徴分布の記述によって★様々な方法が考えられるが、例えば、次式を採用することができる。

【０１５５】

【数２０】

$$G_s(y) = \sum_{i=1}^{M} \beta_i(F_1(y), F_2(y), \cdots, F_M(y), M) \cdot F_i(y) = \sum_{i=1}^{M} \beta_i \cdot F_i(y) \quad \cdots (20)$$

【０１５６】但し、$\beta_i(F_1(y), F_2(y), \cdots, F_M(y), M)$は、ノイズ観測区間Ｔｎの第ｉフレームから得られる無音特徴分布$F_i(y)$に対する重み関数であり、以下、β_iと記述する。なお、重み関数β_iは、次式（２１）の条件を満足するものである。

【０１５７】

【数２１】

$$\sum_{i=1}^{M} \beta_i(F_1(y), F_2(y), \cdots, F_M(y), M) = \sum_{i=1}^{M} \beta_i \equiv 1 \quad \cdots (21)$$

【０１５８】ノイズ観測区間Ｔｎにおける各フレームの特徴ベクトルｙを構成するコンポーネントが無相関であれば、無音特徴分布｛Ｆｉ（ｙ），ｉ＝１，２，・・・，Ｍ｝は、平均ベクトルμ_iと分散マトリクスΣ_iで規定される正規分布Ｎ（μ_i，Σ_i）となる。

【０１５９】この場合、無音モデル補正部３３は、ノイズ観測区間Ｔｎの各フレームから得られる無音特徴分布＊10

$$\mu_{sil} = \frac{a}{M} \sum_{i=1}^{M} \mu_i$$

$$\Sigma_{sil} = \frac{b}{M} \sum_{i=1}^{M} \Sigma_i \quad \cdots (22)$$

【０１６１】ここで、係数ａおよびｂとしては、例えば、シミュレーションにより最適な値を決定することができる。

【０１６２】なお、無音特徴分布Ｆｉ（ｙ）から、無音モデルＧｉ（Ｚ）を生成する方法は、上述の方法に限定されるものではなく、例えば、本件出願人が先に出願した特願2000-276856号（特願平11-375766号を基礎とする国内優先権主張出願）等に開示されている各種の方法を採用することができる。

【０１６３】ところで、上述のように、無音特徴分布を用いて、無音モデルの適応を行う場合においても、特徴ベクトルを用いて、無音モデルの適応を行う場合と同様に、注目音声区間の直前の環境ノイズだけでなく、過去※

$$F_j[I_i] = \begin{bmatrix} f_1(i,j) \\ f_d(i,j) \\ \vdots \\ f_D(i,j) \end{bmatrix} \quad \cdots (23)$$

【０１６６】また、第ｉ発話の音声区間の直前のノイズ観測区間Ｔｎの第ｊフレームの環境ノイズから得られる無音特徴分布Ｆｊ［Ｉｉ］のｄ番目のコンポーネントf_d（ｉ，ｊ）は、上述したことから、式（２４）に示すように、平均値μ_d（ｉ，ｊ）と、分散σ_d^2（ｉ，ｊ）に★

$$f_d(i,j) = N(\mu_d(i,j), \sigma_d^2(i,j)) \quad \cdots (24)$$

【０１６８】この場合、第１の適法方法（図８）では、無音モデル補正部３３は、第１乃至第Ｎ発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズから得られる無音特徴分布のｄ番目のコンポーネントの平均値μ_{sil}　☆

$$\mu_{sil}(d) = \frac{1}{\sum_{i=1}^{N} M(i)} \left[\sum_{i=1}^{N} W_i \cdot \sum_{j=1}^{M(i)} \mu_d(i,j) \right]$$

$$\sigma_{sil}^2(d) = \frac{1}{\sum_{i=1}^{N} M(i)} \left[\sum_{i=1}^{N} W_i \cdot \sum_{j=1}^{M(i)} \sigma_d^2(i,j) \right] \quad \cdots (25)$$

＊Ｆｉ（ｙ）としての正規分布を規定する平均ベクトルμ_iと分散マトリクスΣ_iを用い、例えば、次式にしたがって、無音モデルＧｉ（Ｚ）を表す正規分布を規定する平均ベクトルμ_{sil}と、Σ_{sil}を演算する。

【０１６０】

【数２２】

※の１以上の音声区間の直前の環境ノイズにも基づき、上述の第１乃至第３の３つの適応方法（図８乃至１３）によって、注目音声区間の音声を認識するのに用いる無音モデルの適応を行うことが可能である。

【０１６４】即ち、例えば、いま、第ｉ発話の音声区間の直前のノイズ観測区間Ｔｎの第ｊフレームの環境ノイズから得られる無音特徴分布を、Ｆｊ［Ｉｉ］と表すとすると、本実施の形態では、特徴ベクトルがＤ次のコンポーネントで構成されるから、無音特徴分布Ｆｊ［Ｉｉ］は、次式に示すようなＤ次のコンポーネントで表される。

【０１６５】

【数２３】

★よって規定される正規分布Ｎ（μ_d（ｉ，ｊ），σ_d^2（ｉ，ｊ））で表すことができる。

【０１６７】

【数２４】

☆（ｄ）と分散σ_{sil}^2（ｄ）を、次式にしたがって計算する。

【０１６９】

【数２５】

【０１７０】そして、無音モデル補正部３３は、次式で示される平均ベクトルμ_{sil}と分散マトリクスΣ_{sil}によって規定される正規分布を、注目音声区間の音声の認識＊に用いる無音モデルG_s［I_s］とする。

【０１７１】

【数２６】

$$\mu_{sil}=\begin{bmatrix}\mu_{sil}(1)\\ \vdots\\ \mu_{sil}(d)\\ \vdots\\ \mu_{sil}(D)\end{bmatrix},\quad \Sigma_{sil}=\begin{bmatrix}\sigma^2_{sil}(1) & & 0\\ & \ddots & \\ & & \sigma^2_{sil}(d)\\ & \ddots & \\ 0 & & \sigma^2_{sil}(D)\end{bmatrix} \quad \cdots (26)$$

【０１７２】次に、第２の適応方法（図１０）では、無音モデル補正部３３は、第１乃至第Ｎ－１発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズから得られる無音特徴分布F_1［I_1］乃至F_{s-1}［I_{s-1}］を計算する。さらに、無音モデル補正部３３は、無音特徴分布F_1［I_1］乃至F_s［I_{s-1}］の平均ベクトルμ_{sil-1}と分散マトリクスΣ_{sil-1}を、式（２５）および（２６）に※おける場合と同様に計算し、その平均ベクトルμ_{sil-1}と分散マトリクスΣ_{sil-1}によって規定される正規分布Ｎ（μ_{sil-1}，Σ_{sil-1}）を、式（２７）に示すように、第１の無音モデルG_{s-1}とする。

【０１７３】

【数２７】

$$G_{s-1}=N(\mu_{sil-1},\ \Sigma_{sil-1}) \quad \cdots (27)$$

【０１７４】また、無音モデル補正部３３は、注目フレームである第Ｎ発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズから得られる無音特徴分布F_s［I_s］のｄ番目のコンポーネントについて、そのノイズ観測区★間Ｔｎにおける平均値$\mu_{sil-2}(d)$と分散$\sigma^2_{sil-2}(d)$を、次式にしたがって計算する。

【０１７５】

【数２８】

$$\mu_{sil-2}(d)=\frac{1}{M(N)}\sum_{d=1}^{M(N)}\mu_d(N,j)$$

$$\sigma^2_{sil-2}(d)=\frac{1}{M(N)}\sum_{d=1}^{M(N)}\sigma^2_d(N,j) \quad \cdots (28)$$

【０１７６】さらに、無音モデル補正部３３は、式（２９）で示される平均ベクトルμ_{sil-2}と分散マトリクスΣ_{sil-2}によって規定される正規分布Ｎ（μ_{sil-2}，Σ_{sil-2}）を、式（３０）に示すように、第２の無音モデ☆ルをG_{s-2}とする。

【０１７７】

【数２９】

$$\mu_{sil-2}=\begin{bmatrix}\mu_{sil-2}(1)\\ \vdots\\ \mu_{sil-2}(D)\end{bmatrix},\quad \Sigma_{sil-2}=\begin{bmatrix}\sigma^2_{sil-2}(1) & & 0\\ & \ddots & \\ 0 & & \sigma^2_{sil-2}(D)\end{bmatrix} \quad \cdots (29)$$

【０１７８】

◆ ◆【数３０】

$$G_{s-2}=N(\mu_{sil-2},\ \Sigma_{sil-2}) \quad \cdots (30)$$

【０１７９】以上のようにして、第１の無音モデルG_{s-1}と、第２の無音モデルG_{s-2}を得た後は、無音モデル補正部３３は、第１の無音モデルG_{s-1}と、第２の無音モデルG_{s-2}とを統合することにより、注目音声区間の音声の認識に用いる無音モデルG_s［I_s］を生成する。

【０１８０】即ち、無音モデル補正部３３は、例えば、上述の式（７）にしたがい、第１の無音モデルG_{s-1}を規定する平均ベクトルμ_{sil-1}と、第２の無音モデルG_{s-2}を規定する平均ベクトルμ_{sil-2}とを統合し、平均ベクトルμ_{sii}を求めるとともに、第１の無音モデルG_{s-1}＊を規定する分散マトリクスΣ_{sil-1}と、第２の無音モデルG_{s-2}を規定する分散マトリクスΣ_{sil-2}とを統合し、分散マトリクスΣ_{sii}を求める。そして、無音モデル補正部３３は、その平均ベクトルμ_{sii}と分散マトリクスΣ_{sii}によって規定される正規分布Ｎ（μ_{sii}，Σ_{sii}）を、式（３１）に示すように、注目音声区間の音声の認識に用いる無音モデルG_s［I_s］とする。

【０１８１】

【数３１】

$$G_{sii}=N(\mu_{sii},\ \Sigma_{sii}) \quad \cdots (31)$$

【００８２】次に、第３の適応方法（図１２）では、無音モデル補正部３３は、直前の発話、つまり第Ｎ－１発話の音声区間の音声認識に用いられた無音モデルG_{s}［I_{s-1}］を、音声認識部３２（図１４）から取得する。

【００８３】さらに、無音モデル補正部３３は、注目フレームである第Ｎ発話の音声区間の直前のノイズ観測区間Ｔｎの環境ノイズから得られる無音特徴分布Ｆ［I_{n}］の、そのノイズ観測区間Ｔｎにおける平均ベクトルと分散マトリクスを、上述の式（２８）および（２９）にしたがって計算し、その平均ベクトルと分散マトリクスによって規定される正規分布としての無音モデルG_{n}’［I_{n}］を生成する。

【００８４】そして、無音モデル補正部３３は、第Ｎ－１発話の音声区間の音声認識に用いられた無音モデルG_{s}［I_{s-1}］と、第Ｎ発話の音声区間の直前の環境ノイズだけから得られた無音モデルG_{n}’［I_{n}］とを統合することにより、注目音声区間の音声の認識に用いる無音モデルG_{s}［I_{s}］を生成する。

【００８５】即ち、例えば、第Ｎ－１発話の音声区間の音声認識に用いられた無音モデルG_{s}［I_{s-1}］としての正規分布を規定する平均ベクトルと分散マトリクスを、それぞれμ_{s1-1}　と$\Sigma_{s1\ -1}$とするとともに、第Ｎ発話の音声区間の直前の環境ノイズだけから得られた無音モデルG_{n}’［I_{n}］としての正規分布を規定する平均ベクトルと分散マトリクスを、それぞれμ_{s1-2}　と$\Sigma_{s1\ -2}$とすると、無音モデル補正部３３は、例えば、上述の式（７）にしたがい、平均ベクトルμ_{s1-1}　とμ_{s1-2}　とを統合し、平均ベクトルμ_{s1}を求めるとともに、分散マトリクス$\Sigma_{s1\ -1}$と$\Sigma_{s1\ -2}$とを統合し、分散マトリクスΣ_{s1}を求める。そして、無音モデル補正部３３は、その平均ベクトルμ_{s1}と分散マトリクスΣ_{s1}によって規定される正規分布を、注目音声区間の音声の認識に用いる無音モデルG_{s}［I_{s}］とする。

【００８６】以上のように、特徴分布パラメータを用いる場合においても、注目音声区間の直前の環境ノイズだけでなく、過去の１以上の音声区間の直前の環境ノイズにも基づいて、注目音声区間の音声を認識するのに用いる無音モデルの適応を行うようにすることで、音声区間中の無音部分に十分対処可能な無音モデルの適応を行うことができ、これにより、無音部分に起因する音声認識性能の劣化を防止（低減）することができる。

【００８７】なお、第ｉ発話の音声区間の直前の環境ノイズから得られる無音特徴分布Ｆ［I_{i}］の、そのノイズ観測区間Ｔｎにおける平均ベクトルと分散マトリクスを、上述の式（２８）および（２９）にしたがって計算する場合においては、そのノイズ観測区間Ｔｎにおける Ｍフレームそれぞれから得られる時系列の無音特徴分布F_{1}［I_{1}］，F_{2}［I_{2}］，・・・，F_{4}［I_{4}］は、平等に扱うのではなく、例えば、対応する音声区間に近いほど重みをおいて扱うことが可能である。このような重み付けの方法の詳細については、例えば、上述の特願2000-276856号等に記載されている。

【００８８】次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【００８９】そこで、図１８は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【００９０】プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク１０５やＲＯＭ１０３に予め記録しておくことができる。

【００９１】あるいはまた、プログラムは、フロッピー（登録商標）ディスク、CD-ROM(Compact Disc Read Only Memory)，MO(Magneto optical)ディスク，DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体１１１に、一時的あるいは永続的に格納（記録）しておくことができる。このようなリムーバブル記録媒体１１１は、いわゆるパッケージソフトウエアとして提供することができる。

【００９２】なお、プログラムは、上述したようなリムーバブル記録媒体１１１からコンピュータにインストールする他、ダウンロードサイトから、ディジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(LocaI Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部１０８で受信し、内蔵するハードディスク１０５にインストールすることができる。

【００９３】コンピュータは、CPU(CentraI Processing Unit)１０２を内蔵している。CPU１０２には、バス１０１を介して、入出力インタフェース１１０が接続されており、CPU１０２は、入出力インタフェース１１０を介して、ユーザによって、キーボードや、マウス、マイク等で構成される入力部１０７が操作等されることにより指令が入力されると、それにしたがって、ROM(Read Only Memory)１０３に格納されているプログラムを実行する。あるいは、また、CPU１０２は、ハードディスク１０５に格納されているプログラム、衛星若しくはネットワークから転送され、通信部１０８で受信されてハードディスク１０５にインストールされたプログラム、またはドライブ１０９に装着されたリムーバブル記録媒体１１１から読み出されてハードディスク１０５にインストールされたプログラムを、RAM(Random Access Memory)１０４にロードして実行する。これにより、CPU１０２は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU１０２は、その処理結果を、必要に応

じて、例えば、入出力インタフェース１１０を介して、LCD(Liquid CryStal Display)やスピーカ等で構成される出力部１０６から出力、あるいは、通信部１０８から送信、さらには、ハードディスク１０５に記録等させる。

【０１９４】ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【０１９５】また、プログラムは、１のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【０１９６】なお、本実施の形態では、音響モデルとして、HMMを採用し、HMM法に基づいて音声認識を行うようにしたが、音響モデル、および音声認識のアルゴリズムは、特に限定されるものではない。

【０１９７】また、本発明は、特徴ベクトルと、特徴分布パラメータの両方を用いて音声認識を行う場合にも適用可能である。

【０１９８】

【発明の効果】本発明のモデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置によれば、音声区間の直前の区間で観測される音声データが抽出され、抽出データとして出力される。そして、現在の音声認識の対象となっている音声の音声区間の直前の区間における抽出データと、過去に音声認識の対象とされた音声の音声区間の直前の区間における抽出データとに基づいて、無音を表す音響モデルである無音モデルの適応が行われる。従って、音声区間中の無音部分に十分対処可能な無音モデルの適応を行うことが可能となる。

【図面の簡単な説明】

【図１】本発明を適用した音声認識装置の第１実施の形態の構成例を示すブロック図である。

【図２】図１のノイズ観測区間抽出部３の処理を説明する図である。

【図３】特徴抽出部５の構成例を示すブロック図である。

【図４】音声認識部６の構成例を示すブロック図である。

【図５】HMMを示す図である。

【図６】音声認識処理を説明するフローチャートである。

【図７】各発話に対して、無音モデルの適応を行う様子を示す図である。

【図８】第１の無音モデルの適応方法を説明する図である。

【図９】第１の無音モデルの適応方法による無音モデル適応処理を説明するフローチャートである。

【図１０】第２の無音モデルの適応方法を説明する図である。

【図１１】第２の無音モデルの適応方法による無音モデル適応処理を説明するフローチャートである。

【図１２】第３の無音モデルの適応方法を説明する図である。

【図１３】第３の無音モデルの適応方法による無音モデル適応処理を説明するフローチャートである。

【図１４】本発明を適用した音声認識装置の第２実施の形態の構成例を示すブロック図である。

【図１５】図１４のノイズ観測区間抽出部３の処理を説明する図である。

【図１６】特徴抽出部３１の構成例を示すブロック図である。

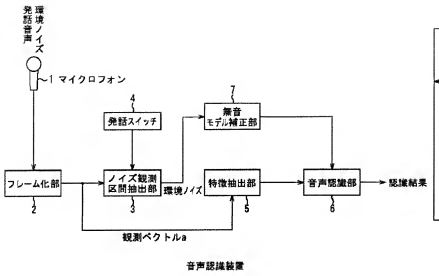
【図１７】音声認識部３２の構成例を示すブロック図である。

【図１８】本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

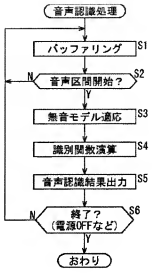
【符号の説明】

１　マイクロフォン，　２　フレーム化部，　３　ノイズ観測区間抽出部，４　発話スイッチ，　５　特徴抽出部，　６　音声認識部，　７　無音モデル補正部，　１１　パワースペクトラム分析部，　２１－１乃至２１－N，２１－ｓ識別関数演算部，　２２　決定部，　３１　特徴抽出部，　３２　音声認識部，　３３　無音モデル補正部，　４２　特徴分布パラメータ算出部，　４３　ノイズ特性算出部，　１０１　バス，　１０２　CPU，　１０３　ROM，　１０４RAM，　１０５　ハードディスク，　１０６　出力部，　１０７　入力部，　１０８　通信部，　１０９　ドライブ，　１１０　入出力インタフェース，　１１１　リムーバブル記録媒体

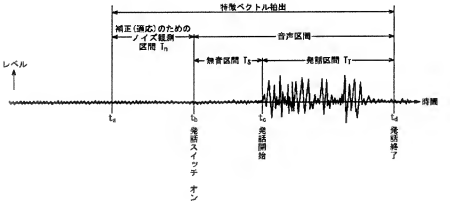
【図１】



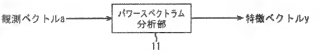
【図６】



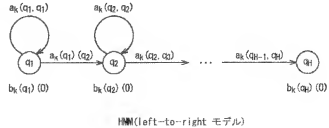
【図２】



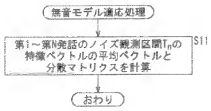
【図３】



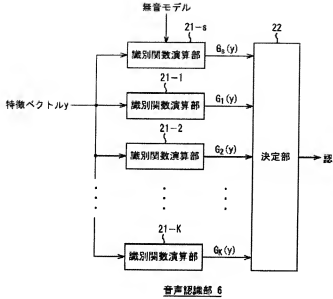
【図５】



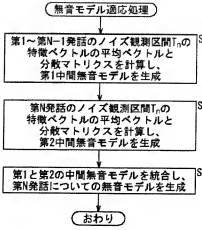
【図９】



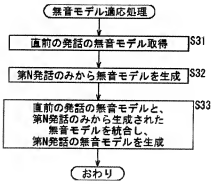
【図４】



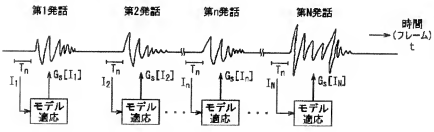
【図１１】



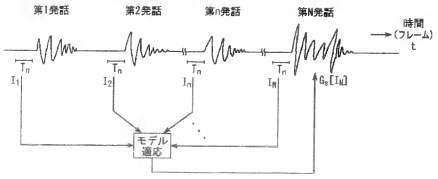
【図１３】



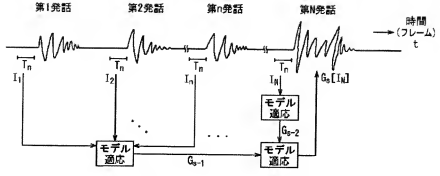
【図７】



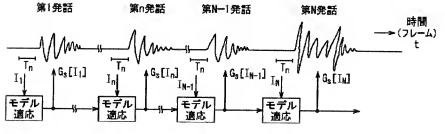
【図８】



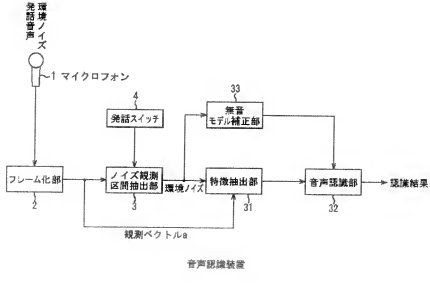
【図１０】


第1発話
第2発話
第n発話
第N発話
時間
(フレーム)
t
Tn
I1
I2
In
IN
GS[IN]
モデル適応
GS-2
GS-1

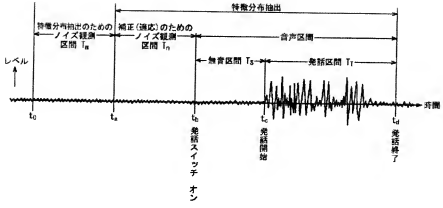
【図１２】


第1発話
第n発話
第N−1発話
第N発話
時間
(フレーム)
t
Tn
I1
GS[I1]
In
GS[In]
IN-1
GS[IN-1]
IN
GS[IN]
モデル適応

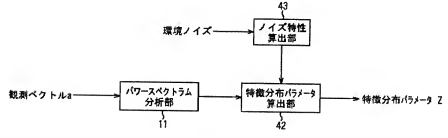
【図１４】


環境ノイズ
発話音声
1 マイクロフォン
4
発話スイッチ
33
無音モデル補正部
フレーム化部
2
ノイズ規測区間抽出部
3
環境ノイズ
特徴抽出部
31
音声認識部
32
認識結果
規測ベクトルa
音声認識装置

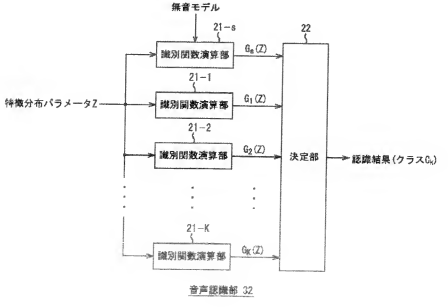
【図１５】


特徴分布抽出
特徴分布用抽出のための
ノイズ経験
区間 T_B
補正（調応）のための
ノイズ経験
区間 T_n
音声区間
無音区間 T_S
発話区間 T_I
レベル
時間
t_0
t_a
t_b
t_c
t_d
発話スイッチ オン
発話開始
発話終了

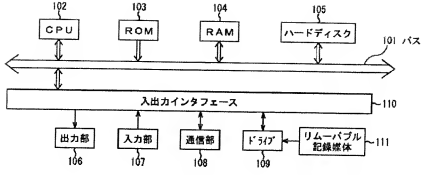
【図１６】


43
環境ノイズ
ノイズ特性
算出部
観測ベクトル a
パワースペクトラム
分析部
11
特徴分布パラメータ
算出部
42
特徴分布パラメータ Z
特徴抽出部 31

【図１７】


無音モデル
21－s
22
識別関数演算部
$G_s(Z)$
21－1
特徴分布パラメータZ
識別関数演算部
$G_1(Z)$
21－2
識別関数演算部
$G_2(Z)$
決定部
認識結果（クラスC_0）
21－K
識別関数演算部
$G_K(Z)$
音声認識部 32

【図１８】



コンピュータ

【手続補正書】
【提出日】平成１２年１１月２８日（２０００．１１．２８）
【手続補正１】
【補正対象書類名】明細書
【補正対象項目名】発明の名称
【補正方法】変更
【補正内容】
【発明の名称】　モデル適応装置およびモデル適応方法、記録媒体、並びに音声認識装置